PI, co-PI: A. Marco Saitta (PR SU), Frédéric Datchi (DR CNRS)

Title: MAchine leaRninG experimental and modeling AppRoaches for Exotic phases of maTter (MARGARET)

Keywords: Liquid-liquid phase transition, exotic phases of matter, critical phenomena, ab initio simulations, diffraction and spectroscopic data analysis and resolution, machine learning, phase space explorations

Scientific context and objective

The liquid-liquid transition (LLT) is a rare and intriguing phenomenon in which a single-component liquid transforms into another one via a first-order transition. Due to their counterintuitive nature, LLTs have intrigued scientists for several years and challenged our perception of the liquid state, for which the notion of polymorphism was long considered impossible. LLTs have been predicted from computer simulations of several systems, and heavily debated in the case of water. So far though, experimental evidence remain scarce and often controversial as they mostly concern supercooled, *i.e.* metastable, liquids where the LLT occurs simultaneously with crystallization, making it hard to separate the two phenomena. A liquid-liquid critical point (LLCP), similar to the well-known gas-liquid critical point, has been predicted at the end of the LLT line in some cases, but until our recent work on sulfur [L. Henry et al, Nature 584 (2020)], never observed in any material. As illustrated in the phase diagram of Fig. 1, taken from the latter work, the LLCP in sulfur terminates a first-order transition line between a low-density liquid (LDL) phase and a high-density liquid (HDL) one.

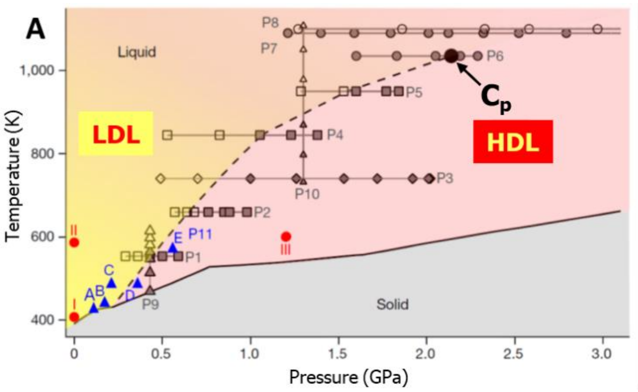

Fig. 1. Phase diagram of sulfur after [L. Henry et al, Nature 584 (2020)]], in the P-T range of the liquid-liquid transition. P1-P11 indicate the various paths along which experimental measurements of density and structure were made.

In the case of water, this putative LLCP has been invoked to understand the numerous thermodynamic anomalies of water in the supercooled regime, but we have recently shown, by theoretical calculations, that water does not exhibit a LLT [under review], possibly ending a 30-year long debate in the community.

Scientific objective

The main objective of this project is to significantly advance our understanding of liquid polymorphism and LLTs, and, by extension, of the liquid state itself, by providing accurate experimental data sets that will constitute a solid basis from which the systematics of LLT can be extracted, and eventually will aid the emergence of theories from which predictions can be made. So far indeed, our understanding of LLT remains quite primitive, and there is no theory able to predict whether a given system will exhibit a LLT. This is why the known experimental realizations remain scarce and have been made rather accidentally. To reach our objective, we propose to study elemental sulfur and phosphorous, over a large range of P-T conditions (0-150 GPa, 300-3000 K) and combine x-ray and optical diagnostics with theoretical studies using state-of-the art ab initio computer simulations and machine learning techniques that will provide information at the microscopic and thermodynamic levels. If time allows, other systems, representative of various types of liquids (network, molecular and metallic liquids), will also be investigated, in parallel with experiments.

Workflow of the PhD project

AI developments for ab initio modelling (supervised by the PI)

We plan to develop those kind of theoretical studies in close collaboration with the experimental tasks, supported by the ANR LILI, whose PI is the co-supervisor of the present PhD project. More specifically, we plan to approach at first the LLT in sulfur and phosphorus. We envisage in particular to gain microscopic insight on the relevant structural mechanisms driving those transitions, and to accurately determine the transition lines and critical points. To this end, it is mandatory to preserve the ab initio level of description of the interatomic interactions. Size effects, however, can be extremely important, and an extensive sampling of the liquid-states configurational spaces is necessary to probe those spaces with statistical significance. This is the major bottleneck from the computational point of view, which will require the development and use of specifically tailored machine-learning-based tools, such as High Dimensional Neural Network Potentials (HDNNPs), in collaboration with J. Behler (Göttingen University), one of the founders of the method, and main developer of the RuNNer code, with whom we have established a fruitful collaboration.

- Neural Networks for interatomic potentials

The first part of the PhD project will be devoted to the development and training of HDNNPs capable to accurately describe the local and global environment of sulfur atoms in the LDL and HDL phases. To this end, nanosecond-long AIMD trajectories, using simulation boxes containing about 100-200 S atoms, will be generated on several points along the isobaric P10 and the isothermal P6 pathways of the above figure. This will allow a first rough determination of the location of the critical point along P6, and to begin to structurally characterize the transformation mechanisms both along P6 and P10. Those configurations will then be used in RuNNer to generate an accurate sulfur HDNNP. The structural proximity of the states along the P6 isobar will certainly improve the reliability and stability of the generated HDNNP. We will need to take into account possible long-range effects in the potential, likely requiring size corrections with respect to the AIMD training slabs. To countercheck the HDNNP results, we will also generate a Gaussian-approximation potential (GAP), a conceptually different machine-learning strategy, using kernel regression rather than NN, that proved effective in the accurate description of other pure elements.

- Data-driven definition of transformation coordinates

In the follow-up we will thus study the LLT in sulfur along the P6 pathway, using the HDNNP and GAP potentials generated as explained above, and our in-house methodology for the efficient statistical sampling of high-dimensional configuration spaces. The latter approach yields free energy landscapes of activated transitions - including possible high barriers - exploiting a combination of state-of-the-art enhanced sampling techniques (metadynamics, committor analysis, umbrella sampling). However, this method needs the definition of efficient one- or very-low-dimensional topological transformation coordinates, yet capable to track nontrivial structural changes in ordered or disordered systems, occurring in a high-dimensional space. If so far in our group we have mostly used physico-chemical intuition to the latter point, we wish in this project to implement more "data-driven" approaches to define the most efficient coordinates of the transformation. In a first work on a textbook chemical reaction, we compare[1] indeed the quality of the textbook reaction coordinate versus the one emerging from a large number of trajectories launched from the transition state. This kind of approach will be generalised during this thesis.

Experimental AI developments (supervised by the co-PI)

In our x-ray diffraction experiments, the measured diffracted signal I(Q) contains that of the sample and that of its environment (pressure cell), and is recorded on a domain of finite wavevectors Q. From this we aim to extract (1) the structure factor S(Q) of the sample and (2) the pair distribution function g(r) from the Fourier transform of S(Q). So far, we have been using an algorithm that automatically optimizes several parameters (scale factor between the measured signal and that of the empty cell, normalization factor), as briefly explained in [L. Henry et al, Nature 584 (2020)]. Other parameters such as the maximum Q (Q_{max}) for the Fourier transformation are also optimized "manually", *i.e.* based on the judgment of the user (we take the highest Q_{max} which does not give

---

[1] T. Magrino, L. Huet, F. Pietrucci, and A. M. Saitta, *Critical assessment of data-driven versus heuristic reaction coordinates in solution chemistry*, to be submitted (2022).

"too much" parasitic oscillation). There is generally no single minimum for these parameters, and the user judgment remains an important "variable". The current project will thus be an ideal playground to implement AI methods capable to make this crucial operation less user-dependent, and even provide an optimized version of this algorithm. To this end, we will use as a test-set at first a simple model system (rare gas) with synthetic data, then on experimental data on liquid sulfur and phosphorous. This part will be supervised by the co-PI.

Another possible route, if time allows, will be to use AI in Reverse Monte Carlo methods, where Monte Carlo simulations are used to obtain a simulated S(Q) for a set of particles that matches the experiment. There are several codes that use various approaches, but generally no single solution and the AI could perhaps help to determine if one or a set of solutions is more likely based on the physics of the system.

Qualification of the PIs

The PI and co-PI have a consolidated expertise and a strong publication record in the field of exotic properties of matter at extreme conditions, including, in the last ten years, 1 Nature, 8 PNAS, 9 Phys Rev Lett, 2 J. Phys. Chem. Lett. 2 Nature Com, 2 Phys Life Rev, 1 ChemSci. As mentioned above, the PI is about to publish one paper using HDNNP, and another one on data-driven definition of reaction coordinates.

General AI impact on the research group and role of the Institute

These planned AI developments are needed to make a major leap in the conceptual, computational, and experimental tools in our research group. Indeed, on the theoretical side, we have a publication, with a 2nd year PhD student, on the study of chemical reactions in solution, using a similar NN approach, about to be submitted[2]. It is however a much more “targeted” problem, from the complexity point of view, than the LLT we aim to study here. In fact, in simple chemical reactions, the number of bonds to be broken and formed is small, and can relatively easily be described by human intuition. On the contrary, collective transformations, as LLT between disordered phases, have too many degrees of freedom to be dealt with through human intuition. On the experimental side, complex experiments at extreme conditions generally provide scarce and “dirty” data to be analysed; the need of automated tools of pattern recognition are clearly necessary to improve the quality and reliability of the information obtained by experiments.

These AI-driven improvement are clearly required both to be able to conduct large-scale predictive and realistic simulations, and to guide and interpret the experiments in this project.

Moreover, they are perfectly in line with the scientific strategy of the group and of the laboratory with respect to AI, and with the role of the PI as director of the recently installed GDR “*Intelligence Artificielle en science des MATériaux (IAMAT)*”.

Finally, SCAI clearly is the right place to carry out this project, as the PhD candidate will clearly take advantage of the favourable environment, including schools and tutorials, to learn about and carry out the needed AI developments.

Identified candidate for the PhD fellowship

We have already identified a Master 2 student having the required strong background in statistical physics, computational materials science, condensed matter physics and chemistry.

Her name is Sonia Salomoni, student on a double Master diploma (University of Bologna and Sorbonne Université) in Materials Science, and after the 1st semester of M2 at SU she is the major of a promotion of more than 40 students. She is a Master2 intern in our group right now, and she is willing to undertake these innovative methods and determined to carry out this project within a strong collaboration between theory and experiments.

---

[2] T. Devergne, T. Magrino, F. Pietrucci, and A. M. Saitta, *Combining machine learning approaches and accurate ab initio enhanced sampling methods for prebiotic chemical reactions in solution*, to be submitted (2022).